

知识管理论坛
2016 年 第 6 期(总第 6 期)

【学术探索】

基于多特征融合的跨域情感分类模型研究

◎ 琚春华[1,2] 邹江波[1,3] 傅小康[2]

[1] 浙江工商大学管理科学与电子商务学院 杭州 310018
[2] 浙江工商大学现代商贸研究中心 杭州 310000
[3] 浙江工商大学管理学院 杭州 310018

摘要：[目的 / 意义] 跨领域情感分类仍是亟需重点研究的问题之一。[方法 / 过程] 借助情感无关词，通过谱聚类算法构建源领域与目标领域的跨域情感特征词簇，将谱聚类得到的情感词特征与位置特征、关键词特征、词性特征融入逻辑回归分类算法中，实现基于多特征融合的跨领域情感分类算法；并以用户评论数据进行验证。[结果 / 结论] 研究结果表明，CDFF(Cross Domain pulse Four Factor) 算法可有效实现跨域用户的情感分类，为跨领域情感分类研究提供借鉴。

关键词：跨域情感分类　多特征融合　谱聚类　迁移学习
分类号：TP391

引用格式：琚春华 , 邹江波 , 傅小康 . 基于多特征融合的跨域情感分类模型研究 [J/OL]. 知识管理论坛 , 2016, 1(6): 464-470[引用日期]. http://www.kmf.ac.cn/p/1/83/.

1 引言

互联网用户的交互行为产生了大量评论数据，如客户购买某商品后的评论、微博用户针对热点话题的评论等。这些交互数据中隐含着用户对某类事物的情感倾向，它对构建用户兴趣模型、产生推荐结果具有重要实践意义。情感分类即根据评论数据将用户情感分为两类：积极和消极，虽然人们可以很容易从某条评论数据中推测出当时评论者的情感，但对于机器来说并非易事，并且一些评论数据中并没有显性的表达出用户情感，这更增大了机器学习的难度。

国内外已有许多学者通过半监督学习的方法对情感分类问题进行了研究[1-3]，有研究者为了更好地利用关键句和细节句之间的差异性和互补性，将抽取的关键句分别用于有监督和半监督的情感分类中[2]，但如何准确判断出评论的关键句仍是需要继续深入研究的问题。有研究者使用大规模未标记数据和少量情绪词实现了情感分类[3]，虽然降低了人工标记数据的成本，但模型不能重复在其他领域中使用，仍需针对特定领域进行情感分类学习。在情感分类研究中也有针对如何计算情感词的情感度，有

基金项目：本文系国家自然科学基金资助项目“电商环境下融入在线社会关系的消费信贷价值度量研究”（项目编号：71571162）和浙江省自然科学基金资助项目“融入物联情境的商业数据流挖掘模型及可靠性研究”（项目编号：LY14F020002）研究成果之一。
作者简介：琚春华，院长，教授，博士；邹江波（OCIRD: 0000-0002-2811-0004），实验员，博士研究生, E-mail: zjgsu518@163.com；傅小康，讲师，博士。
收稿日期：2016-10-16　　发表日期：2016-12-30　　本文责任编辑：徐健

知识管理论坛，2016(6):464-470
DOI：10.13266/j.issn.2095-5472.2016.054

学者针对情感词的情感度确定问题进行研究[4]，提出了模糊层次分析法来度量情感词的情感度。这些方法的分类结果依赖于手工标识的训练数据，训练数据好的分类准确率也高，但实际情况却是每个领域中手工标识形成分类训练数据的代价是很高的，如果对每个领域都进行手工数据标识也是不现实的，因此有研究者考虑到情感分类任务的领域相关性[5]，通过跨领域学习减少情感分类的数据标记，提出一种基于评价对象类别的跨领域学习方法，但评价对象类别粒度较粗，不适合跨多个领域的情感分类[6]。由此可见，在某一个领域情感训练产生的分类准确的分类器未必能在另一个领域中表现出同样的准确性。为了解决情感分类算法领域依赖性高、人工数据标记成本大等问题，本文对跨域情感分类进行了深入研究，发现通过谱聚类可缩短不同领域间情感词的距离，在已有研究的基础上，本文希望借助情感无关词来桥接源领域与目标领域，再利用谱聚类算法将不同领域的情感词聚集到一起，并考虑相关特征进行融合，以此实现跨领域情感分类。

2 概念定义与问题描述

本节对领域、情感词、跨域情感分类等相关概念做出了相关定义。

定义 1 领域：一个领域 D 代表现实世界中一类实体或概念的集合。

可理解为超市中不同的产品区域，有食品、文具、家电等，图书馆中不同学科领域，领域的粒度可抽象或细分，具体需根据实际情况而定。

定义 2 情感词：给定一个特定的领域，情感词是那些能够反映用户情感倾向的词语。

这些情感词与用户短语表达出来，通过语句拆分可组成情感词序列 $[w_1,w_2,w_3…w_n]$，本研究中没有考虑情感词在语句的排序对最终情感分类的影响，但考虑了情感词在语句中的位置对最终情感分类的影响，每个特定的领域 D 有属于本领域的情感词库 W（w_iÎW），借鉴 bag-of-words 的思想，将 $c(w_i,x_j)$ 表示为情感词 w_i 在语句 x_j 中出现的频率。

定义 3 情感分类：给定领域，根据语句 x_i 整体语义表达划分情感类别 y_i（正面 y_i=1 或负面 y_i=-1）将已标记情感类别的语句组成情感分类中的训练数据 (x_i,y_i)，将未被标记情感类别的语句称为预测数据。

定义 4 跨域情感分类：给定两个不同的领域，源领域（D_{src}）和目标领域（D_{tar}），假定源领域中含已标记数据集（$[x_{srci},y_{srci}]$,iÎ1,2…n_{src}），目标数据集含未标记数据集（$[x_{tarj}]$,jÎ1,2…n_{tar}），如果某个分类器能通过在源领域训练学习准确预测目标领域中未标记的数据集，那么将这样的分类称为跨域情感分类。

跨域情感分类需要解决领域依赖的问题，即相邻领域情感词的表达是相近的，而实际情况中，用户通常会针对不同的领域发表与领域相关的评论语，如表 1 列举了新浪微博中用户对电影和社会两大类别中相关热点微话题的评论，用户评语短语显性或隐性地表达了评论主体的某些情感，由此看出用户对当前话题的情感倾向，具有情感倾向的情感词已在表中用黑体标出，如正面情感词“激动”“激烈”“给力”等，负面情感词“痛苦”“折磨”等。但每个领域中的情感词却存在区别，如电影领域中的负面情感词 “俗套”“凌乱”等，社会领域中正面情感词“合理” 等，其中的“俗套”“凌乱”“合理”属于领域相关词，“既然”“毕竟”属于领域无关词。

除此之外，位置特征、关键词、词性特征也是情感分类中需要考虑的问题，一般评论语句的最后几个情感特征最能表达评论者的情感，其次，如果出现如“但是”“毕竟”“我认为”等转折关键词，评论者的情感表达可能发生转变，最后，大多数能表达用户情感的都是形容词或副词，因此在情感分类时，除情感特征外，也需要考虑上述特征因素对情感分类的影响。

因此，结合国内外相关研究，给出了跨领域的情感分类框架，如图 1 所示：

表 1 跨域情感词对比

类别	电影	社会
负面情感	《盗墓笔记》记得刚听说要拍的时候，各种幻想各种猜，激动得很，现在真说在准备的时候，全剩担心了，毕竟以前好多小说拍出来的效果是达不到书面文字的	《高考英语改革》数学原来和英语一样让我恶心啊，虽然好多年不碰了，这一碰又想起那些年被英语和数学痛苦折磨的我，而如今又被一些奇葩的数学憋出了内伤了，憋的差点死了！
正面情感	《金刚狼 2》五场战役激烈紧凑，感受狼叔健硕的肌肉大战恶势力，4D 视觉效果更给力。	《星巴克咖啡价格》怎么能这么算呢？不能仅仅算咖啡的成本，还有店面租金、人工、水电、税等等，既然他定这么高的价格而且还有人去消费，那就是合理的！
负面情感	《特殊身份》再喜欢的演员都掩盖不了我对此片的差评。开场画面就已经让人深切地感受到了 20 世纪 90 年代香港警匪片的浓烈汗臭。剧情俗套不说，打斗场面真心也毫无亮点，甄子丹已经过度消费了武戏。画面同样感到凌乱。	近日哈尔滨雾霾，哈尔滨在试供热和气象因素双重作用下，空气质量直线下降，连续出现灰霾天。昨日，哈市 12 个监测点中 8 处重度污染，其中两个监测点 PM2.5 浓度超过 300 毫克每立方米，达到严重污染。这是今年秋末冬初首次爆出的最严重污染天。出门记得戴口罩！

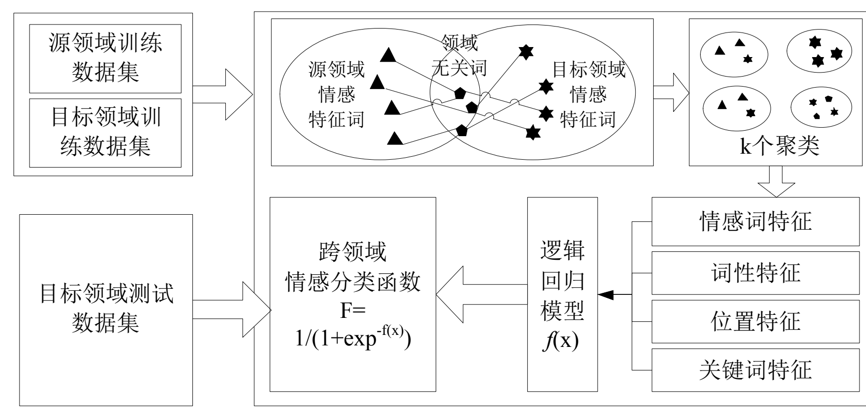


图 1 跨领域情感分类框架

其中目标领域情感特征词根据标识数据获得，但实际情况是该标识数据量较少或没有现成的标识数据，或需要人工标识部分数据。借助情感无关词，通过谱聚类算法构建了源领域与目标领域的跨域情感特征词簇，将谱聚类得到的情感词特征与位置特征、关键词特征、词性特征等 4 种因子融入逻辑回归分类算法中，实现基于多特征融合的跨领域情感分类算法。

3 跨域情感分类模型

本文借鉴了林政等基于情感关键句抽取的情感分类方法 [2]，但不是为了抽取关键句，而是将文献中的特征得分用于最终情感分类，考虑了情感特征（即领域情感词）、位置特征、关键词特征及词性特征，其中的情感特征通过多领域谱聚类得到，词性特征剔除与情感分类无关的词，以此达到跨领域情感分类的目的。因此，考虑上述 4 个特征的情感分类可用公式(1)表示，此时每一条评论数据共 4 属性特征，都是通过计算特征的得分进行分类，削弱了特征空间对跨领域分类的影响。p_0 是偏置项，p_1、p_2、p_3、p_4 参数，它们可通过训练数据训练得出。

$$f(x_i) = p_0 + p_1 \times f_{sentimental(x_i)} + p_2 \times f_{position(x_i)} + p_3 \times f_{keywords(x_i)} + p_4 \times f_{speech(x_i)} \quad \text{公式（1）}$$

通过公式（1）计算的值并不能表达情感分类（正面或负面），因此加入公式（2），达到对跨领域情感分类的目的。

$$F = \frac{1}{1+e^{-f(x_i)}} \quad \text{公式（2）}$$

此时函数 ø 将 $f(x_i)$ 的值域映射到 0 和 1 上，这样便可达到情感分类的目的。

3.1 情感特征词

通过评论短语的情感特征词大体可以判断

知识管理论坛，2016(6):464–470
DOI：10.13266/j.issn.2095–5472.2016.054

评论者的情感倾向，通常在情感分类时情感特征词的权重较大，跨领域分类遇到的关键问题就是不同领域中的情感特征空间不同，最终导致源领域训练的分类器不能很好的应用到目标领域中。因此本文借助领域无关词作为桥梁[6]，通过谱聚类方法实现跨领域的情感词转换，得到新的情感词特征空间，在该空间里通过公式（3）计算评论短语 x_i 的情感特征词的得分。

$$f_{sentimental(x_i)}=\frac{\sum_{j=1}^{n}\mathrm{positive}(w_{ij})-\sum_{j=1}^{n}\mathrm{negative}(w_{ij})}{n}$$

公式（3）

每条评论短语 x_i 需要分词并剔除停顿词，其中 $\mathrm{positive}(w_{ij})$ 表示第 i 条评论语句的第 j 个词在谱聚类集中对应情感词，该词在聚类中代表正面情感特征倾向；$\mathrm{negative}(w_{ij})$ 表示第 i 条评论语句的第 j 个词在谱聚类集中对应情感词，该词在聚类中代表负面情感特征倾向，n 是该评论短语中剔除停顿词后的总词数。

3.2 词性特征

词性特征属于领域无关的特征，虽然每个领域都有其特定的特征空间，但这些特征空间的词性都是相同的，有文献指出形容词和副词往往最能代表了跨领域评论的情感倾向[7]，而名词则和领域相关，因此考虑目标领域的词性特征进行情感分类，参照 B. Pang 等学者的方法[1]首先对评论短语进行 POS 词性标注；再按照预定义的规则抽取目标领域评论短语中的形容词和副词词语；最后使用公式 (4) 计算每条评论短语的词性比重得分。

$$f_{\mathrm{speech}(x_i)}=\frac{w_i}{n}$$

公式（4）

其中 w_i 等于按照预定义规则抽取的第 i 条评论短语中形容词和副词词语总数，n 等于第 i 条评论短语中提出评论短语后的总词数。该公式表示形容词和副词在评论短语中所占的比重，即形容词和副词对情感分类的影响程度。

3.3 位置特征

一条评论语句中可能包含多个正面情感词和负面情感词，但通常最可能表达评论者情感的情感词出现在评论的开始或结尾，需考虑情感评论中的位置特征对情感分类的影响，因此，位置特征的得分可通过公式（5）计算得出。

$$f_{\mathrm{postion}(x_i)}=\sum_{j=1}^{M}\alpha*\mathrm{pos}(w_{ij})^2+\beta*\mathrm{pos}(w_{ij})+c$$

公式（5）

$\mathrm{pos}(w_{ij})$ 表示第 j 个词在第 i 条评论语句中的位置，可看出位置特征服从一元二次函数，即抛物线图像，以此达到凸显句前和句尾词在情感分类中重要性的目的，但也不能与中间位置差异过大，因此抛物线的开口应该大，防止两端值对情感分类的影响过大。

其中满足：

$$-\frac{\beta}{2\alpha}=\frac{M}{2};\alpha>0;\beta>0$$

M 表示 x_i 中的总字数，中间位置是函数的最低点，此处计算的情感词得分较低，而位于评论开头和结尾的情感词得分较高。由于针对短评数据，句中特征词数据较少，难以判断情感分类，此时位置特征的影响力削弱，可适当调整 c 的取值，调整位置特征的得分。

3.4 关键词特征

情感分类中，评价短语中的关键词能够反映出评论这情感倾向的变化，因此需考虑关键词特征对情感倾向性的影响，本文归纳了多领域中的 20 个常用关键词用于实验，这些情感词包括：总之、我认为、然而、毕竟、但是、既然等，这里不再一一列出。关键词特征的计算如公式（6）所示：

$$f_{kewords(x_i)}=\sum_{j=1}^{n}keyword(w_{ij})$$

公式（6）

其中：

$$\mathrm{keyword}(w_{ij})=\begin{cases}=1 w_{ij}\in\mathrm{keyword}\\=0 w_{ij}\notin\mathrm{keyword}\end{cases}$$

3.5 基于多特征融合的跨域情感分类算法

为了实现跨领域情感分类，本算法除了通过谱聚类算法将情感词特征空间进行映射以外，还融入了词性特征、位置特征、关键词特征，在新的特征空间上训练得到逻辑回归分类器，具体算法步骤如下：

算法 1：基于多特征融合的跨域情感分类算法

输入：源领域训练数据，少量目标领域训练数据，聚类个数 k；

输入：逻辑回归分类器。

算法步骤：

步骤（1） 剔除训练数据集停顿词；

步骤（2） 针对源领域训练数据和少量目标领域训练数据采用谱聚类算法得到 k 个聚类；

步骤（3） 根据谱聚类结果通过公式（3）计算训练数据集的情感特征词的得分；

步骤（4） 通过公式（4）计算词性特征得分；

步骤（5） 通过公式（5）计算训练数据集的位置特征得分；

步骤（6） 根据关键词词典通过公式（6）计算训练数据集的关键词特征得分；

步骤（7） 对训练数据集进行词性标注，抽取训练数据集中的副词和形容词；

步骤（8） 将训练数据集进行转换，以情感词、位置、关键词、词性、情感为特征，构建新的训练数据集 D_{new}；

步骤（9） 根据新的训练数据集通过梯度下降法学习得到公式（1）中参数 p_0，p_1，p_2，p_3，p_4 的值；

步骤（10）将参数带入公式（2）输出逻辑回归分类器。

算法 2：谱聚类算法 [8]：

输入：源领域训练数据，目标领域训练数据，聚类个数 k；

输出：k 个聚类。算法步骤：

步骤（1） 根据领域无关和领域相关词语构造双向图 G（$V_{DS} \cup V_{DI}$, E），计算图双向图的带权领接矩阵 $W \in R^{n\times n}$，如果 $i \neq j$，$W_{ij}=m_{ij}$，否则 $W_{ij}=0$；

步骤（2） 计算对角矩阵 D，其中 $D_{ii}=\Sigma_j W_{ij}$，构建图的拉普拉斯矩阵 $L=D^{-1/2}WD^{-1/2}$；

步骤（3） 计算拉普拉斯矩阵 L 的前 k 个最大特征值对应的特征向量并构建成特征矩阵 $U=[u_1, u_2 \ldots u_k] \epsilon R^{n\times k}$；

步骤（4） 标准化特征矩阵 U，

$$U_{ij}=\frac{U_{ij}}{(\sum_j U_{ij}^2)^{1/2}};$$

步骤（5） 在矩阵 U 上使用 K-means 算法，将 n 个点聚类到 k 个聚类中；

步骤（6） 返回 k 个聚类。

4 实验分析与结果

4.1 实验设置

为了验证模型的有效性，本文采用 Java 语言，基于 weka 的逻辑回归源代码实现了算法 CDFF。针对了数据集，采用中国科学院计算技术研究所的分词软件接口 ICTCLAS（http://ictclas.org）和开源项目 IKAnalyzer，加入了搜狗实验室中的互联网词库（http://www.sogou.com/labs/resources.html）和本文搜集整理的停顿词典，实现了对文本进行分词及词性附加操；SVM 算法使用的是标准工具包 light-SVM（http://svmlight.joachims.org）采用线性核函数；通过谱聚类算法实现跨领域情感词的转换，由于情感特征的得分依赖于聚簇，因此实验中会调整聚类参数 k 的值来比较跨领域情感分类的效果。

4.2 实验结果与分析

本文用到的数据集来自网络用户对酒店、电脑（笔记本）与书籍 3 个领域的短评平衡数据（http://www.searchforum.org.cn/tansongbo/corpus-senti.htm），其中每个领域的正负类各 2 000 篇，共 12 000 条平衡评论数据，数据集的具体组成如表 2 所示：

知识管理论坛，2016(6):464–470
DOI：10.13266/j.issn.2095–5472.2016.054

表 2 数据集描述 （单位：条）

数据名称	正面评论	负面评论	平均长度
酒店	2 000	2 000	118
电脑 (笔记本)	2 000	2 000	87
书籍	2 000	2 000	102

数据集上领域的相关度并不是很大，为了验证本算法的有效性，采用 6 个跨领域情感分类任务方案：酒店→电脑，酒店→书籍，电脑→酒店，电脑→书籍，书籍→酒店，书籍→电脑；其中箭头前表示源领域，箭头后表示目标领域。采用支持向量机（SVM）、SFA（Spectral Feature Alignment）、SCL（Structural Correspondence Learning）[13]3 种算法与本文算法 CDFF 作对比，针对每个算法的实验都采用五折交叉验证，即随机划分每一领域数据为 5 份，每次取其中 4 份进行训练，一份进行测试，然后把 5 次分类结果的平均结果作为最终结果。

考虑到谱聚类中聚簇的个数会影响情感特征词的得分，因此在实验中分别设置簇的个数为 5、10、15 来度量其对情感分类的影响。具体如表 3 所示：

表 3 跨领域分类结果

所跨领域	SVM	SCL	SFA	CDFF（k=5）	CDFF（k=10）	CDFF（k=15）
酒店→电脑	0.692 1	0.724 8	**0.749 1**	0.693 2	0.732 6	0.745 8
酒店→书籍	0.719 3	0.741 5	0.731 3	0.724 3	**0.793 1**	0.748 8
电脑→酒店	0.692 3	0.749 1	**0.761 0**	0.703 5	0.755 3	0.735 0
电脑→书籍	0.762 1	0.798 0	0.811 3	0.782 7	**0.829 8**	0.817 7
书籍→酒店	0.663 2	0.669 4	0.731 7	0.724 2	**0.793 8**	0.781 7
书籍→电脑	0.765 3	0.806 8	0.812 4	0.818 2	**0.829 7**	0.813 9
跨领域分类准确率平均值	0.715 7	0.748 3	0.766 1	0.741 0	0.782 4	0.773 8

从表 3 的跨领域平均准确值中可以看出本算法的实验结果较 SFA 算法高，高出情感分类的准确率随聚簇的个数增加而增加，但当 k=15 时，准确率增加的效果已不明显，但从 5 个簇到 10 个簇时，分类准确率提高，由此可见谱聚类个数会影响跨领域情感分类的结果。

本算法除考虑情感特征词外还加入了位置特征、关键词特征、词性特征，为了验证加入这些特征的有效性，通过固定聚簇的个数（k=10），逐次加入这些特征后对比算法准确性，来观察不同特征对跨领域情感分类的影响，具体如表 4 所示：

表 4 依次加入相关特征后的跨领域情感分类准确率

加入特征	酒店→电脑	酒店→书籍	电脑→酒店	电脑→书籍	书籍→酒店	书籍→电脑
加入情感词特征	0.684 3	0.738 1	0.732 7	0.810 9	0.725 1	0.782 6
加入词性特征	0.686 4	0.745 1	0.755 3	0.819 9	0.733 8	0.797 4
加入位置特征	0.708 9	0.749 6	0.785 6	0.820 3	0.747 8	0.813 3
加入关键词特征	0.739 3	0.750 2	0.799 0	0.825 8	0.775 2	0.827 7

从表 4 中可以看出依次分别加入词性特征、位置特征、关键词特征后跨领域情感分类的准确率均有所提高，但是每个特征的贡献率不同，从表 4 中可看出，位置特征和关键特征的贡献率平均大于词性特征的贡献率。因此通过上述两个实验验证了基于多特征融合的跨领域分类算法可提高情感分类准确率。

5 总结与展望

虽然人们可以很容易的从某条评论数据中推测出当时评论者的情感，但对于机器来说并非易事，本文借助情感无关词搭建源领域与目标领域的桥梁，通过谱聚类算法将不同领域的情感词聚集到一起，应用谱聚得到的特征集计



算目标领域测试数据的情感得分，与传统谱聚类算法不同，本文在跨领域情感分类时还考虑了位置特征、词性特征、关键词特征对最终情感分类的影响，将谱聚类得到聚类中的特征与位置、词性、关键词特征融合以此实现跨领域情感分类。通过在用户评论数据上对本算法进行实验，验证了本算法在跨域用户情感分类时的有效性。由于本文选择的数据集较为标准，但微博评论数据中存在很大的随意性，领域相关词也比较新颖，因此针对微博数据特性的跨领域情感分类将是未来研究的重点。

参考文献：

[1] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[EB/OL].[2015-10-12].http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf.

[2] 林政 , 谭松波 , 程学旗 . 基于情感关键句抽取的情感分类研究 [J]. 计算机研究与发展 , 2012, 9(11): 2376-2382.

[3] 代大明，王中卿，李寿山 , 等 . 基于情绪词的非监督中文情感分类方法研究 [J]. 中文信息学报 , 2012, 26(4): 103-108.

[4] 李纲 , 王忠义 , 寇广增 . 情感分类中情感词的情感倾向度的计算方法研究 [J]. 情报学报 , 2011, 28(3): 292-298.

[5] 张慧 , 李寿山 , 李培峰 , 等 . 基于评价对象类别的跨领域情感分类方法研究 [J]. 计算机科学 , 2013,40(1): 229-233.

[6] PAN S J, NI X C, SUN J T, et al. Cross-domain sentiment classification via spectral feature alignment[EB/OL].[2015-10-18].https://www.microsoft.com/en-us/research/wp-content/uploads/2010/04/Cross-Domain-Sentiment-Classification-via-Spectral-Feature-Alignment.pdf.

[7] RUI X, CHENG Q Z. A POS-based ensemble model for cross-domain sentiment classification[EB/OL].[2015-10-26].https://www.researchgate.net/publication/228841203_A_POS-based_Ensemble_Model_for_Cross-domain_Sentiment_Classification.

[8] 张志武 . 跨领域迁移学习产品评论情感分析 [J]. 现代图书情报技术 , 2013(6): 49-54.

[9] 马凤闸 , 吴江宁 , 杨光飞 . 基于双重选择策略的跨领域情感倾向性分析 [J]. 情报学报 , 2012, 31(11): 1202-1209.

[10] 张迪 . 基于跨领域分类学习的产品评论情感分析 [D]. 上海：上海交通大学 , 2010.

[11] DANUSHKA B, DAVID W, JOHN C. Cross-domain sentiment classification using a sentiment sensitive thesaurus[J]. IEEE transactions on knowledge and data engineering, 2013, 25(8): 1719-1731.

[12] TAN S B,CHENG X Q, GHANEM M M,et al. A novel refinement approach for text categorization[EB/OL].[2015-11-02]. http://dl.acm.org/citation.cfm?id=1099554.1099687.

[13] BLITZER J, DREDZE M, PEREIRA F. Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification[EB/OL].[2015-11-11].http://www.cs.jhu.edu/~mdredze/publications/sentiment_acl07.pdf.

作者贡献说明：

琚春华：提出基于多特征融合的跨域情感分类模型，论文撰写、修改；

邹江波：实现基于多特征融合的跨域情感分类模型算法，论文撰写、修改、定稿；

傅小康：参与模型提出和算法实现，论文撰写、修改。

Cross-domain Emotion Classification Model Based on the Multi-feature Fusion

Ju Chunhua[1,2] Zou Jiangbo[1,3] Fu Xiaokang[2]

[1]School of Management Science & E-commerce, Zhejiang Gongshang University, Hangzhou 310018

[2]Center for Studies of Modern Business, Hangzhou 310000

[3]School of Business Administration,, Zhejiang Gongshang University, Hangzhou 310018

Abstract: [Purpose/significance] The sentiment classification is still one of the cross-cutting issues needed to focused on. **[Method/process]** With the help of emotion unrelated words, by the spectral clustering algorithm, the authors constructed a cross-domain feature words emotion cluster in the source and target areas of the field. The position of the features and characteristics of emotional words, keyword features, and POS features were integrated into the logic of the regression classification algorithm to achieve a cross-cutting emotion classification algorithm based on the multi-feature fusion. **[Result/conclusion]** Research results show that CDFF (Cross-domain pulse Four Factors) algorithm is effective when the cross-domain user emotion is classified and its provide reference for same study.

Keywords: cross-domain sentiment classification   multi-feature fusion   spectral clustering   transfer learning